CONSTRAINED DISCRIMINATIVE PLDA TRAINING FOR SPEAKER VERIFICATION

Johan Rohdin, Sangeeta Biswas and Koichi Shinoda

Department of Computer Science, Tokyo Institute of Technology, Tokyo, Japan {johan, sangeeta}@ks.cs.titech.ac.jp, shinoda@cs.titech.ac.jp

ABSTRACT

Many studies have proven the effectiveness of discriminative training for speaker verification based on probabilistic linear discriminative analysis (PLDA) with i-vectors as features. Most of them directly optimize the log-likelihood ratio score function of the PLDA model instead of explicitly train the PLDA model. But this optimization process removes some of the constraints that normally are imposed on the PLDA log likelihood ratio score function. This may deteriorate the verification performance when the amount of training data is limited. In this paper, we first show two constraints which the score function should follow, and then we propose a new *constrained* discriminative training algorithm which keeps these constraints. Our experiments show that our method obtained significant improvements in the verification performance in the male trials of the telephone speaker verification tasks of NIST SRE08 and SRE10.

Index Terms— PLDA, discriminative training, speaker verification, i-vector

1. INTRODUCTION

In recent years, the combination of i-vector [1], [2] and probabilistic linear discriminant analysis (PLDA) [3], [4] has become the stateof-the-art system in speaker verification. In this system, an i-vector *extractor*, maps utterances into low dimensional vectors known as i-vectors. The i-vectors contains information related to speaker identity as well as irrelevant factors such as the transmission channel or the speakers' emotion. Given two i-vectors, i.e., one from the *enrollment* phase and one from the *test* phase, the PLDA model separates speaker factors from irrelevant factors and provides a log-likelihood ratio (LLR) score for the two i-vectors being from the same speaker or not.

In this study we point out two interesting properties of the PLDA LLR score function. The first, and most interesting, is a *directional property*; the score will always be higher for two i-vectors pointing in the same direction than for two i-vectors pointing in the opposite direction. Since the effectiveness of cosine similarity, (e.g., [2] and [5]), indicates that speakers are well discriminated by angular information, this property seems desirable. The second is a *length property*; the score for two equal i-vectors will be reduced if one of them is scaled with a factor α and the other with factor $1/\alpha$.

In the original work for face recognition, [3], as well as in many works in speaker verification (e.g., [4] and [6]), the PLDA model parameters have been optimized under the Maximum likelihood (ML) criteria.

In [7] and [8], discriminative training of the PLDA model parameters were shown to outperform ML training in speaker verification. Their proposed discriminative training scheme optimizes the log-likelihood ratio score function of the PLDA model directly, instead of training the PLDA model explicitly. This optimization process allows the score function to be more general than the score function of a standard (ML trained) PLDA model. As a consequence, the above mentioned properties of the PLDA LLR may not be preserved.

Obviously, given sufficiently large amounts of training data, a more general model can be expected to perform better than a constrained model. But training a PLDA model already requires very large amounts of training data. In addition, it has been shown that discriminative training of probabilistic models needs more training data than ML training [9]. Adding extra flexibility to the model may therefore be more harmful than useful when the amount of training data is limited. This taken into account, keeping the good properties of PLDA model and use them as constraints in the discriminative training, may improve the verification performance.

In this study, we propose a new *constrained* discriminative training algorithm which keeps some of the properties of the PLDA scoring function. Experimentally, we show that our method obtained significant improvements in the verification performance in the male trials of the telephone speaker verification tasks of NIST SRE08 and SRE10.

2. I-VECTOR AND PLDA BASED SPEAKER VERIFICATION

2.1. Log-likelihood ratio (LLR) score

In the i-vector system [2], it is assumed that a *Gaussian Mixture Model* (GMM) *-supervector*, μ , corresponding to an utterance can be modeled as

$$\boldsymbol{\mu} = \bar{\boldsymbol{\mu}} + \boldsymbol{T}\boldsymbol{\omega},\tag{1}$$

where $\boldsymbol{\omega}$ is a random vector known as the *i-vector*, \boldsymbol{T} is a basis matrix for the *total variability space*, i.e., speaker and channel variability, of $\boldsymbol{\mu}$, and $\bar{\boldsymbol{\mu}}$ is the mean of $\boldsymbol{\mu}$. It is assumed that $\boldsymbol{\omega}$ follows standard normal distribution and that its dimension, d, i.e., the rank of \boldsymbol{T} , is lower than dimension of $\bar{\boldsymbol{\mu}}$.

In [4], it was proposed to use PLDA in speaker verification with i-vectors as features. In that study, a modification of the original PLDA model, suitable for low-dimensional features was suggested. It models i-vectors, ω , as

$$\boldsymbol{\omega} = \boldsymbol{m} + \boldsymbol{V}\boldsymbol{y} + \boldsymbol{D}\boldsymbol{z}, \tag{2}$$

where y and z are random vectors depending on the speaker and session respectively. The speaker variability is given by V and the channel variability is given by D. The elements of y and z are assumed to be independent with standard normal distribution. Usually, rank(V) < d but rank(D) = d gives the best performance. For the special case when rank(V) = d, the model is referred to as the *twocovariance* model [10].

For scoring two i-vectors, ω_i and ω_j , we need to calculate the log-likelihood ratio of the hypothesis that the two i-vectors are from

the same speaker, \mathcal{H}_s , and the hypothesis that they are from different speakers, \mathcal{H}_d , i.e.,

$$s_{ij} = \log \frac{p(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j | \mathcal{H}_s)}{p(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j | \mathcal{H}_d)}$$

= $\log \frac{\int P(\boldsymbol{\omega}_i | \boldsymbol{y}) P(\boldsymbol{\omega}_j | \boldsymbol{y}) P(\boldsymbol{y}) d\boldsymbol{y}}{\int \int P(\boldsymbol{\omega}_i | \boldsymbol{y}_1) P(\boldsymbol{\omega}_j | \boldsymbol{y}_2) P(\boldsymbol{y}_1) P(\boldsymbol{y}_2) d\boldsymbol{y}_1 d\boldsymbol{y}_2},$ (3)

since the speaker factors, y, are the same if the two i-vectors are from the same speaker. Eq. (3) has a closed form solution. It is given by:

$$s_{ij} = \boldsymbol{\omega}_i^T \boldsymbol{P} \boldsymbol{\omega}_j + \boldsymbol{\omega}_j^T \boldsymbol{P} \boldsymbol{\omega}_i + \boldsymbol{\omega}_i^T \boldsymbol{Q} \boldsymbol{\omega}_i + \boldsymbol{\omega}_j^T \boldsymbol{Q} \boldsymbol{\omega}_j + (\boldsymbol{\omega}_i + \boldsymbol{\omega}_j)^T \boldsymbol{c} + \boldsymbol{k},$$
(4)

where

$$\boldsymbol{P} = \frac{1}{2} \Sigma_{\text{tot}}^{-1} \Sigma_{\text{ac}} (\Sigma_{\text{tot}} - \Sigma_{\text{ac}} \Sigma_{\text{tot}}^{-1} \Sigma_{\text{ac}})^{-1}$$
(5)

$$\boldsymbol{Q} = \frac{1}{2} \Sigma_{\text{tot}}^{-1} - (\Sigma_{\text{tot}} - \Sigma_{\text{ac}} \Sigma_{\text{tot}}^{-1} \Sigma_{\text{ac}})^{-1}$$
(6)

$$\boldsymbol{c} = -2(\boldsymbol{P} + \boldsymbol{Q})\boldsymbol{m} \tag{7}$$

$$k = \frac{1}{2} (\log |\Sigma_{\text{tot}}| - \log |\Sigma_{\text{tot}} - \Sigma_{\text{ac}} \Sigma_{\text{tot}}^{-1} \Sigma_{\text{ac}}|) + \boldsymbol{m}^T 2(\boldsymbol{P} + \boldsymbol{Q}) \boldsymbol{m},$$
(8)

and $\Sigma_{ac} = VV^T$ and $\Sigma_{tot} = VV^T + DD^T$.

2.2. Constraints on the PLDA LLR score function

P and Q are symmetric and have the same rank as V [6]. In addition, it can be shown based on Eq. (5) and (6), that the matrices, P and Q, are constrained as follows:

- 1. **P** is positive-definite
- 2. Q is negative-definite

In the above, *definite* needs to be replace with *semidefinite* when the rank of V is lower than d. Notice that these constraints are not sufficient to keep all the original properties of a PLDA model.

2.3. Properties of the PLDA LLR score

In this subsection we show that the above mentioned constraints on P and Q gives the score function in Eq. (4) some interesting properties. The first constraint leads to a *directional property*. Consider an i-vector, ω , scored against both $\alpha \omega$ and $-\alpha \omega$, where α is a positive constant. That is, in the first trial, ω is scored against an i-vector pointing in same direction and, in the second trial it is scored against an i-vectors pointing in the opposite direction. If the i-vectors are centered around m, the difference between the scores given by Eq. (4) of these two trials is

$$s(\boldsymbol{\omega}, \alpha \boldsymbol{\omega}) - s(\boldsymbol{\omega}, -\alpha \boldsymbol{\omega}) = 2\alpha \boldsymbol{\omega}^T \boldsymbol{P} \boldsymbol{\omega}.$$
 (9)

In other words, the score of the *same direction* trial will be guaranteed to be larger than the score of the *different direction* trial if and only if *P* is positive definite.

The second constraint leads to a *length property*:

$$s(\boldsymbol{\omega}, \boldsymbol{\omega}) > s(\alpha \boldsymbol{\omega}, \frac{1}{\alpha} \boldsymbol{\omega})$$
 (10)

This property means that two i-vectors of equal length and direction will obtain a higher score than two i-vectors having just equal direction. At a first glance, this may not seem to be a useful property when the i-vectors are length-normalized, but notice that after i-vectors have been length normalized, their mean is not necessarily zero so that centered around this mean, their lengths are not always equal to 1.

3. DISCRIMINATIVE PLDA TRAINING

In [3] and [4] the parameters m, V and D were trained by the ML criteria. Instead of using the ML criteria for training the PLDA model, we can use discriminative training, which directly optimizes the model for discriminating between the same speaker trial and the different speaker trial. This was first proposed in [7] and [8]. In those studies, the parameters, P, Q, c and k, of the scoring function in Eq. (4), were trained directly instead of the parameters, m, V and D of the PLDA model in Eq. (2). Let $t_{ij} \in [-1, 1]$ be the label of the trial (ω_i, ω_j), i.e., t_{ij} equals 1 if ω_i and ω_j are from the same speaker and -1 if they are not from the same speaker. Further, let $\theta = \text{vec}([P, Q, c, k])$, where $\text{vec}(\cdot)$ stacks the columns of a matrix into a column vector. Then θ can be trained discriminatively by minimizing the total loss

$$E(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{j=1}^{n} l\left(t_{ij}, s_{ij}(\boldsymbol{\theta})\right) + R(\boldsymbol{\theta}), \qquad (11)$$

where l(t, s) is a *loss function* for a trial, n is the number of i-vectors in the training set and $R(\theta)$ is a regularization term. That is, we are minimizing the sum of the loss for all possible i-vector pairs in the training set. In this study we will follow [7] and use the logistic regression loss function given by

$$l(t_{ij}, s_{ij}) = \log(1 + \exp(-t_{ij}s_{ij})).$$
(12)

Then, the gradient of in $E(\theta)$ in (11) is given by, ([8]),

$$\nabla E(\boldsymbol{\theta}) = \begin{bmatrix} \nabla_{\boldsymbol{P}} E(\boldsymbol{\theta}) \\ \nabla_{\boldsymbol{Q}} E(\boldsymbol{\theta}) \\ \nabla_{\boldsymbol{c}} E(\boldsymbol{\theta}) \\ \nabla_{\boldsymbol{k}} E(\boldsymbol{\theta}) \end{bmatrix}$$
$$= \begin{bmatrix} 2 \operatorname{vec}(\boldsymbol{\Omega} \boldsymbol{G} \boldsymbol{\Omega}^{T}) \\ 2 \operatorname{vec}([\boldsymbol{\Omega} \circ (\mathbf{1}_{A} \boldsymbol{G})] \boldsymbol{\Omega}^{T}) \\ 2 [\boldsymbol{\Omega} \circ (\mathbf{1}_{A} \boldsymbol{G}) \boldsymbol{\Omega}] \mathbf{1}_{B} \\ \mathbf{1}_{B}^{T} \boldsymbol{G} \mathbf{1}_{B} \end{bmatrix}.$$
(13)

where $\mathbf{1}_A$ is a $d \times n$ matrix of ones and $\mathbf{1}_B$ is a $n \times 1$ matrix of ones, $\mathbf{\Omega} = [\boldsymbol{\omega}_1 \dots \boldsymbol{\omega}_n], \circ$ denotes the element wise multiplication of two matrices and

$$G_{ij} = \frac{\partial l(t_{ij}, s_{ij})}{\partial s_{i,j}}.$$
(14)

Since, in this approach, the score function is optimized directly without any constraints, the obtained parameters may not preserve the properties of the PLDA LLR score discussed in Section 2.2. We will refer to this discriminative training method as *unconstrained discriminative training*.

An alternative approach for discriminative PLDA training was presented in [11]. In that study, a generalization of the PLDA model was considered in order to deal with multiple enrollment sessions. In the case of only one enrollment session, the method corresponds to standard PLDA. In their discriminative training scheme, only the eigenvalues of the covariance matrices of the PLDA model, or, a scaling factor of them was trained discriminatively while remaining parameters were obtained from ML training. As long as the eigenvalues are kept positive, the resulting parameters will preserve the properties of the PLDA model. But since the orientation of the eigenvectors are unchanged, the method does not take full advantage of discriminative training.

4. CONSTRAINED DISCRIMINATIVE PLDA TRAINING

As described in the previous section, the previous studies on discriminative PLDA training either do not preserve the properties of the PLDA model or they do not do discriminative training of all its parameters. In this section we will describe how P and Q can be constrained to be positive- and negative-(semi)definite in discriminative training of the score function in Eq. (4). Moreover, we will constrain the rank of P and Q to be lower than d, since, for ML trained models, this typically performs better than rank equal to d. We call this training *constrained discriminative training*.

The matrix \boldsymbol{P} is positive-semidefinite if

$$\boldsymbol{P} = \boldsymbol{P}_A \boldsymbol{P}_A^T, \tag{15}$$

where P_A is a $d \times r$ matrix with real elements. Accordingly, in order to keep P positive-semidefinite, we train P_A instead of P. The rank of P is equal to r and can therefore be selected by selecting the number of columns in P_A . Based on Eq. (13), the gradient of Eq. (11) with respect to P_A is given by

$$\left[\nabla_{\boldsymbol{P}_{A}} E\right] = \left[4 \operatorname{vec}(\boldsymbol{\Omega} \boldsymbol{G} \boldsymbol{\Omega}^{T} \boldsymbol{P}_{A})\right].$$
(16)

The complexity of this calculation is $\mathcal{O}(n^2d + d^2(n+r))$ compared to $\mathcal{O}(n^2d + d^2n)$ for the unconstrained training [21]. Since $n \gg r$, the additional computational cost is small.

In order to prevent over-fitting, we use L2 regularization towards the ML model, $R(\boldsymbol{\theta}) = \rho \|\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\|^2$ where $(\bar{\boldsymbol{\varphi}})$ refers to the ML estimate. The regularization term for \boldsymbol{P} is $\rho \|\boldsymbol{P}_A \boldsymbol{P}_A^T - \bar{\boldsymbol{P}}\|^2$. The gradient is given by,

$$\left[\nabla_{\boldsymbol{P}_{A}}E\right] = 4\rho \left[\operatorname{vec}\left(\left[(\boldsymbol{P}_{A}\boldsymbol{P}_{A}^{T}) - \bar{\boldsymbol{P}}\right]\boldsymbol{P}_{A}\right)\right].$$
(17)

In order to keep Q negative-semidefinite, we set $Q = -Q_A Q_A^T$ and train Q_A instead of Q. The gradient for Q_A is obtained by a similar modification of Eq. (13) as for P_A . Using these gradients, we then minimize total loss in Eq. (11) with respect to $\theta_A =$ vec($[P_A, Q_A, c, k]$) with the L-BFGS algorithm [12].

5. EXPERIMENTS

In this section we experimentally compare ML training, the *unconstrained* discriminative training and our proposed *constrained* discriminative training.

5.1. Experimental setup

We used NIST SRE 2006 core task as the development set and NIST SRE 2008 core condition-6 (tel-tel) and NIST SRE 2010, coreextcoreext condition-5 (tel-tel) as the evaluation sets. We used EER and the *old* and *new* MDC as evaluation metrics. See [13]-[15] for details. The development set was used to select the regularization parameter, ρ , that minimized old MDC.

Voice activity detection using spectral subtraction [16] was used for removing non speech. For features we used 15 PLP coefficients and log-energy plus their first-order and second-order derivatives.We applied feature warping [17] before applying VAD. We used genderdependent systems. For training the UBM and i-vector extractor, we used NIST SRE 2004 and 2005, Switchboard II-Phase 1, 2 and 3, Switchboard Cellular -Part 1 and 2. The dimension of the i-vector, *d*, was set to 400. For PLDA training we used the same sets except

Table 1. Male results, MDC refers to old MDC for sre06 and sre08-6 and new MDC for sre10-5.

Set	Training method	MDC	EER
Sre06 (Dev)	ML	0.0116	2.24
	Unconstrained Discr.	0.0122	2.28
	Constrained Discr.	0.0110	2.06
Sre08-6	ML	0.0250	4.87
	Discr. Unconstrained	0.0256	4.66
	Discr. Constrained	0.0231	4.62
Sre10-5	ML	0.000378	1.96
	Unconstrained Discr.	0.000417	2.26
	Constrained Discr.	0.000391	1.96

Table 2. Female results, MDC refers to old MDC for sre06 and sre08-6 and new MDC for sre10-5.

Set	Method	MDC	EER
Sre06 (Dev)	ML	0.0181	3.38
	Unconstrained Discr.	0.0156	3.16
	Constrained Discr.	0.0145	3.09
Sre08-6	ML	0.0284	5.89
	Unconstrained Discr.	0.0262	5.63
	Constrained Discr.	0.0265	5.56
Sre10-5	ML	0.000483	3.03
	Unconstrained Discr.	0.000515	3.05
	Constrained Discr.	0.000525	3.08

Switchboard II-Phase 1. The number of i-vectors in the training data was 12383 for female and 9152 for male.

The i-vectors were whitened, i.e., normalized with the total covariance, and length-normalized [6] prior to PLDA training. The rank of V was set to 250.

ML training was performed with the EM algorithm as described in [18]. Discriminative training was started from the ML model, using eigendecomposition to get initial P_A and Q_A . For optimization, we used the L-BGFS method in [19]. We used its default stopping criteria and in addition we stopped the training if no change in old MDC had been observed on the development set for 20 iterations.

5.2. Results

The MDC and EER for male are shown in Table 1. The DET curves are shown in Figs. 1 and 2. For sre08-6, the constrained discriminative training performs better than both ML and unconstrained discriminative training. For sre10-5, ML training is clearly better than the unconstrained discriminative training but our constrained discriminative training gives similar results to ML training except for a small advantage of ML training in the new MDC. As can be seen in Fig. 2, this advantage is only in the very left of the DET curve.

For female, the differences between the methods were smaller. The EER and MDC are shown in Table 2. The DET curve for sre08-6 is given in Fig. 3. Constrained and unconstrained training give very similar results but ML training is clearly worse except for new MDC. For sre10-5, all training methods give comparable results.

Overall, there is no clear winner between ML training and the unconstrained discriminative training proposed in the previous studies. Constrained discriminative training on the other hand, is always



Fig. 1. Det curve for Sre08-6, Male



Fig. 2. Det curve for Sre10-5, Male

better or comparable to the best of the other two methods on all the data sets evaluated in this study. We also confirmed that the matrices P and Q obtained by the unconstrained discriminative training indeed did not fulfill the two constraints of discussed in Subsection 2.2. The violation of the directional property may therefore be one of the causes for the worse performance of the unconstrained discriminative training.

For the female evaluation sets, there were no significant difference between the constrained and unconstrained discriminative training. This may partly be explained by the fact the female training data were larger since, with more training data, there should be



Fig. 3. Det curve for Sre08-6, Female

less need for constraints.

On an Intel Xeon 5670, the CPU-time was typically 2 to 6 hours for the unconstrained training depending on the value of ρ . Despite the slightly higher complexity of the gradient calculation, the training time was reduced by approximately half for the constrained training. This is because the optimization converged in fewer iterations.

6. CONCLUSION AND FUTURE WORK

We presented two properties of the PLDA scoring function that should be preserved in discriminative PLDA training. And then, we developed a novel constrained discriminative training method for preserving them. Experimentally, we show that it outperforms unconstrained discriminative PLDA training.

Future work will include experiments with smaller amounts of training data. Further, in addition to the constraints presented in this study, it can be shown that, for example, P + Q should be positive definite. It would be interesting to evaluate this constraint also. In this study, we used regularization towards the ML model but other approaches should be evaluated, including techniques discussed in [20].

Finally, since the extra flexibility of the unconstrained discriminative training is not helpful, one might ask what kind of extra flexibility could benefit the PLDA model. In [21], it was shown that the PLDA scoring function can be seen as the second order Taylor expansion of a general scoring function. Considering higher order Taylor expansions with suitable constraints could be a good approach.

7. REFERENCES

[1] N. Dehak, R. Dehak, P. Kenny, N. Brümmer, P. Ouellet, and P. Dumouchel, "Support vector machines versus fast scoring in the low-dimensional total variability space for speaker verification," in *Proc. INTERSPEECH*, pp. 1559-1562, 2009.

- [2] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 4, pp. 788-798, 2011.
- [3] S. J. D. Prince and J. H. Elder, "Probabilistic linear discriminant analysis for inferences about identity," in *ICCV*, pp. 1-8, 2007.
- [4] P. Kenny, "Bayesian speaker verification with heavy tailed priors," in *Proc. Odessey*, 2010.
- [5] S. Shum, N. Dehak, E. Chuangsuwanich, D. A. Reynolds, and J. R. Glass, "Exploiting intra-conversation variability for speaker diarization.," in *Proc. INTERSPEECH*, pp. 945-948, 2011.
- [6] D. Garcia-Romero and C. Y. Espy-Wilson, "Analysis of ivector length normalization in speaker recognition systems," in *Proc. INTERSPEECH*, pp. 249-252, 2011.
- [7] L. Burget, O. Plchot, S. Cumani, O. Glembek P., Matejka N., and Brümmer, "Discriminatively trained probabilistic linear discriminant analysis for speaker verification," in *Proc. ICASSP*, pp. 4832-4835, 2011.
- [8] S. Cumani, N. Brümmer, L. Burget, and P. Laface, "Fast discriminative speaker verification in the i-vector space," in *Proc. ICASSP*, pp. 4852-4855, 2011.
- [9] A. Y. Ng and M. I. Jordan, in NIPS, pp. 841-848, 2001.
- [10] N. Brümmer and E. de Villiers, "The speaker partitioning problem," in *Proc. Odessey*, pp. 194-201, 2010.
- [11] B. J. Borgström and A. McCree, "Discriminatively trained bayesian speaker comparison of i-vectors," in *Proc. ICASSP*, pp. 7659-7662, 2013.
- [12] D. C. Liu and J. Nocedal, "On the limited memory bfgs method for large scale optimization," *Mathematical Programming*, vol. 45, pp. 503-528, 1989.
- [13] "NIST SRE06 evaluation plan," http://www.itl. nist.gov/iad/mig/tests/spk/2006/sre-06_ evalplan-v9.pdf.
- [14] "NIST SRE08 evaluation plan," http://www.itl. nist.gov/iad/mig/tests/sre/2008/sre08_ evalplan_release4.pdf.
- [15] "NIST SRE10 evaluation plan," http://www.itl.nist. gov/iad/mig/tests/sre/2010/NIST_SRE10_ evalplan.r6.pdf.
- [16] M. W. Mak and H. B. Yu, "Robust voice activity detection for interview speech in nist speaker recognition evaluation," in *Proc. APSIPA ASC*, 2010.
- [17] J. Pelecanos and S. Sridharan, "Feature warping for robust speaker verification," in *Proc. Odyssey*, pp. 213-218, 2010.
- [18] N. Brümmer, "Em for probabilistic Ida," https://sites. google.com/site/nikobrummer, 2010.
- [19] "minFunc.m," http://www.di.ens.fr/~mschmidt/ Software/minFunc.html.
- [20] C. M. Bishop and J. Lasserre, "Generative or Discriminative? Getting the Best of Both Worlds," *BAYESIAN STATISTICS*, vol. 8, pp. 3–24, 2007.
- [21] S. Cumani, N. Brümmer, L. Burget, P. Laface, O. Plchot, and V. Vasilakakis, "Pairwise discriminative speaker verification in the i-vector space," *IEEE Transactions on Audio, Speech & Language Processing*, vol. 21, no. 6, pp. 1217-1227, 2013.